Hierarchical Clustering of Global COVID-19 Statistics: Comparative Insights from Pandemic Indicators

Noor Razzaq Abbas[1,*], Ghassan AL-Thabhawee[2], Isam Bahaa Aldallal[3], Mostafa Abotaleb[4], Klodian Dhoska[5]

[1]Al-Furat Al-Awsat Technical University, Technical Institute of Najaf, Najaf, Iraq
[2]Sciences of Mathematics, Computer Sciences, College of Health and Medical Techniques-Kufa, Al-Furat Al-Awsat Technical University, Kufa, Iraq
[3]Department of Electrical and Computer Engineering, Altinbas University, Istanbul, Turkey
[4]Engineering School of Digital Technologies, Yugra State University, Khanty-Mansiysk, 628012, Russia
[5]Department of Production and Management, Polytechnic University of Tirana, 1001, Tirana, Albania
Emails: noor.hachame@atu.edu.iq; gmohammed@atu.edu.iq; isam.aldallal@gmail.com; abotalebmostafa@bk.ru; kdhoska@fim.edu.al

Abstract

Hierarchical clustering is applied in this research to study world COVID-19 data up to January 2025 and partition the primary clusters of countries based on epidemiological criteria. Total cases, deaths, recoveries, active cases, tests, population, and per-million were the data explored and were standardized and thereafter analyzed employing agglomerative hierarchical clustering with Ward linkage. The assessment yielded an average Silhouette of 38.5%, Davies–Bouldin value of 0.87, and Calinski–Harabasz value of 77.6, reflecting cluster validity in separation.

The application of dendrograms and PCA projections to plot identified four clusters, reflecting differences in the severity of COVID-19 impacts and responses. Clustering analysis revealed that the high-burden clusters accounted for almost 45% of global death, while low-burden clusters were predominant in over 40% of nations with fewer than 100,000 accumulated instances. The outcomes illustrate hierarchical clustering as an unsupervised learning approach to analyzing epidemiological data and give quantitative estimates to facilitate comparative public health interventions across communities.

Keywords: Hierarchical Clustering; COVID-19; Agglomerative Clustering; Silhouette Score; PCA; Epidemiology; Machine Learning; Unsupervised Learning; Cluster Analysis; Global Health

1. Introduction

The COVID-19 pandemic has triggered widescale research on clustering algorithms in modeling patterns of transmission, heterogeneity at the country level, and epidemiological dynamics. Various studies have proved the capabilities of unsupervised learning, especially clustering algorithms, to distinguish between country-level patterns and predict severity of outbreaks.

James and Menzies [14] explored United States' COVID-19 pathways using the second surge pattern through chaos theory and highlighting the significance of time series modelling in identifying epidemic wave patterns. Similarly, Zarikas et al. [15] conducted clustering analysis on COVID-19 country-specific case numbers and divided nations based on epidemiological values to analyze patterns of outbreaks similarities. Their study proved that clustering could reveal hidden patterns in large amounts of data to facilitate comparative examination of country reaction.

DOI: https://doi.org/10.54216/JAIM.100202
Received: April 02, 2025 Revised: June 28, 2025 Accepted: September 06, 2025

Other clustering techniques have also been investigated. Mahmoudi et al. [16] used a fuzzy clustering technique in comparing the spread rate of COVID-19 among risky countries. They demonstrated in their study that fuzzy techniques could account for uncertainties and overlapping membership of clusters as per the simplicity of disease transmission. Pasin [17] focused on country clustering, i.e., cases and deaths, and found cluster comparisons helpful to examine mortality and morbidity patterns at the international level. Ramadan et al. [18] employed the multivariate data analysis approach based on statistics delivered daily to increase knowledge of changing pandemic parameters for countries. Such findings make it reasonable to utilize several variables in cluster models to explain scale and time development.

Time series clustering has been a matter of key concern.

James and Menzies [19] suggested a two-way evolution model with clusters for multivariate time series COVID-19 analysis. It explained how epidemic curves develop in a temporal manner over countries and find similar patterns in pandemics' waves. They are used mainly to understand how interventions work over regions and terminate how peak outbreaks synchronize.

In addition to pandemic response measures, a few studies have investigated the structural determinants of responses. Kavanagh et al. [20] underscored the importance of democracy, capacity, and coercion as comparative COVID-19 response determinants and contended that political institutions and arrangements of governance shape intervention designs and epidemiological outcomes. Global Health Security (GHS) Index [21] and Oppenheim et al.'s epidemic preparedness index [22] demonstrate how clustering analysis can be enhanced by global preparedness measures. Investigators can identify clusters both by outcome and system readiness using preparedness measures combined with epidemiologic data. Clustering analysis is guided by clinical and epidemiologic studies. Wiersinga et al. [23] provided a critical synopsis of COVID-19 pathophysiology, transmission, diagnosis, and treatment, highlighting the necessity of considering its multifactorial ethology. Richardson et al. [24] published presenting features and outcomes among 5,700 New York City hospitalised patients with identification of comorbidities and risk factors which directly influenced mortality rates. Similarly, so too have Weinberger et al. [25] quantified excess COVID-19-attributable mortality in the US during March and May of 2020 to determine the long-term societal cost of the pandemic. Excess death and clinical information are critical to the elucidation of explanation of clusters insofar as they are explanatory variables for why some country groups carry heavier burdens of mortality than do other country groups.

In aggregate, networked work demonstrates that COVID-19 data clustering analysis is a useful tool towards the identification of epidemiological patterns and the classification of countries based on disease burden, diffusion patterns, and policy responses. From fuzzy clustering [16] to time series-based approaches [19], these approaches supplement traditional epidemiological models by providing unsupervised information regarding the patterns of similarity and contrast of the pandemic impact. The addition of preparedness indicators [21], [22] and clinical factors [23]-[25] further aids in cluster outcome interpretation, enabling further comprehensive analysis of global health vulnerabilities and strengths.

2. Related Work

The COVID-19 pandemic has triggered rigorous examination of clustering techniques to delineate patterns of spread, national level heterogeneity, and epidemiological dynamics. A number of studies have demonstrated that unsupervised learning, in the form of clustering techniques, can clearly differentiate among country trends and forecast outbreak severity.

James and Menzies [14] examined trends of COVID-19 in the United States based on the second surge pattern through chaos theory and with an emphasis on the importance of time series modeling for the discovery of epidemic wave patterns. Zarikas et al. [15] applied cluster analysis to group countries based on COVID-19 cases and epidemiological variables to research overlapping patterns in outbreak trends. Their research found that clustering was able to find latent patterns in large sets of data, allowing comparative study of country-level responses.

Various approaches to clustering have been explored also.

Mahmoudi et al. [16] suggested a fuzzy clustering approach to compare rates of COVID-19 transmission for high-risk countries.

Their research asserted that fuzzy approaches can capture uncertainty and overlap of membership across clusters, consistent with the complexity of disease transmission. Pasin [17] focused on country

DOI: https://doi.org/10.54216/JAIM.100202
Received: April 02, 2025 Revised: June 28, 2025 Accepted: September 06, 2025

clustering, i.e., deaths and cases, and deduced that cluster comparison is necessary for global mortality and morbidity pattern analysis. Ramadan et al. [18] adopted a multivariate data analysis strategy based on daily statistics to enhance the understanding of the changing pandemic indicators from nations. These findings validate the application of multiple variables in cluster models to represent scale and temporal variation. Clustering of time series has been an important problem.

James and Menzies [19] suggested a two-way model of evolution based on clusters for COVID-19 multivariate time series analysis. It enabled us to decipher the trends in epidemic curves across nations over time and quantify corresponding trends of pandemic waves. It is mainly utilized to identify the efficacy of interventions across locations and to examine synchronization of outbreak peaks. In addition to pandemic response program criticism, other research has addressed the structure of responses. Kavanagh et al. [20] emphasized the role of democracy, capability, and coercion in determining relative COVID-19 responses, theorizing that political institutions and regimes affect intervention tactics and epidemiologic outcomes.

The Global Health Security (GHS) Index [21] and Oppenheim et al.'s pandemic preparedness index [22] illustrate the way in which global indicators of readiness can complement clustering analysis.

Researchers are able to specify clusters based on outcomes and system-level preparedness with the help of readiness markers and epidemiological data.

Clinical and epidemiological research are the bedrock of clustering analysis.

Wiersinga et al. [23] performed an in-depth analysis of COVID-19 evolution, transmission, diagnosis, and treatment, giving utmost significance to its identification as having a heterogeneous origin. Richardson et al. [24] documented the clinical course and prognosis of 5,700 New York City hospital patients, showing comorbidities and risk factors that have immediate implications on fatality rates. Weinberger et al. [25] quantified excess mortality caused by COVID-19 from March to May 2020 in the US to clarify the lingering social effects of the pandemic. Death and clinical data justify the clusters, and they serve as explanation variables for variation in death burden between country groupings. The collaborative assessment of the clusters of COVID-19 data shows that this strategy is very effective in identifying epidemiological patterns and categorizing countries based on illness burden, transmission, and policy actions. The techniques, from fuzzy clustering [16] to time series methods [19], complement traditional models of epidemiology through providing unsupervised insight into the patterns of similarity and dissimilarity of the impacts of the pandemic.

The inclusion of preparedness indices [21], [22] and clinical indicators [23]-[25] increases the interpretability of cluster results, thus enabling more focused investigation of global health deficiencies and strengths.

3. Data and Methodology

A. Dataset

The data set used in this study was obtained from Kaggle, titled Global COVID-19 Statistics (Jan 2025) [26]. It records cumulative COVID-19 data for 231 countries/regions up to January 2025. The data set comprises 21 attributes, which represent both absolute and relative pandemic metrics. Some of the prominent attributes are total confirmed cases, total deaths, total recoveries, active cases, serious/critical cases, and testing numbers. To make comparisons across nations of varying population sizes comparable, per-million population metrics such as total cases per million, deaths per million, active cases per million, and tests per million are included. Population size and continental group are also added as contextual factors. Preprocessing such as missing value treatment, numeric attribute normalization, and formatting inconsistency correction was required for the dataset before analysis. These actions ensured that hierarchical cluster models were developed on reliable and similar indicators of different countries.

Figure 1 shows the procedure adopted to pre-process the COVID-19 dataset before hierarchical clustering analysis. Data collection begins, where the raw dataset was obtained from Kaggle and represents global COVID-19 data up to January 2025. During data cleansing, inconsistencies such as string-formatted numbers, commas, and special characters were removed, with all variables being re-

DOI: https://doi.org/10.54216/JAIM.100202
Received: April 02, 2025 Revised: June 28, 2025 Accepted: September 06, 2025

coded into numeric form. Finally, missing values were addressed using median imputation to achieve completeness with the least amount of bias. It was followed by feature selection for the inclusion of nine prominent epidemiological indicators, including total cases, deaths, recoveries, active cases, and per-million metrics, as well as population.

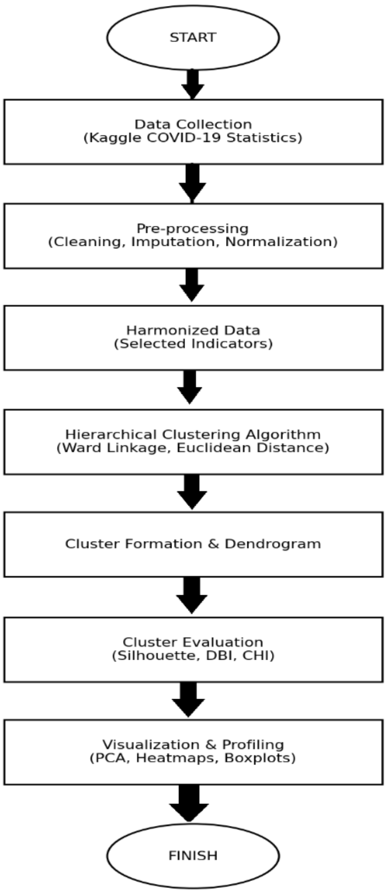


Figure 1: Data Pre-processing.

B. Hierarchical clustering

Hierarchical clustering is a popular unsupervised machine learning technique that organizes data points into a hierarchy of nested clusters according to measures of similarity. Let the nearest be defined as:

$$D = \{x_1, x_2, \ldots, x_n\}, x_i \in R^m \tag{1}$$

where n is the number of countries or samples and m is the number of epidemiological variables. The algorithm begins by taking each observation as its own cluster and then combines clusters sequentially until all of the dataset is in a single cluster. The simple concept is to minimize intra-cluster distance and maximize inter-cluster separation.

The similarity between two points x_i and x_j can be computed using the Euclidean distance:

$$d(x_i, x_j) = \sqrt{\sum_{f=1}^{m} (x_{if} - x_{ff})^2} \tag{2}$$

which is the most commonly applied measure in epidemiological datasets. Other alternatives include the Manhattan distance:

$$d(x_i, x_j) = \sum_{f=1}^{m} |x_{if} - x_{jf}| \tag{3}$$

and cosine similarity:

$$\mathrm{sim}(x_i, x_j) = \frac{x_i \cdot x_j}{\|x_i\|\|x_j\|} \tag{4}$$

depending on the structure of the data.

Clusters are merged according to linkage criteria that define how distances between sets of points are measured. In single linkage the definition is:

$$d(A,B) = \min_{x \in A, y \subset B} d(x,y) \tag{5}$$

while in complete linkage the maximum distance is used:

$$d(A,B) = \max_{x \in A, y \subset B} d(x,y) \tag{6}$$

Average linkage computes the mean of pairwise distances:

$$d(A,B) = \frac{1}{|A||B|} \sum_{x \subset A} \sum_{y \in B} d(x,y) \tag{7}$$

to balance between extremes. Ward's linkage criterion instead seeks to minimize the variance within clusters, expressed as:

$$\Delta(A,B) = \frac{|A||B|}{|A| + |B|} \|\bar{x}_A - \bar{x}_B\|^2 \tag{8}$$

where $\bar{x}_A$ and $\bar{x}_B$ are cluster centroids.

The entire sequence of merges is recorded in a dendrogram where branch heights correspond to linkage distances, providing a graphical representation of hierarchical structure.

To obtain a fixed number of clusters, the dendrogram is cut at a threshold τ such that:

$$C = \{C_1, C_2, \dots, C_k\}, d(C_i, C_j) > \tau \tag{9}$$

for any two distinct clusters [27]-[34].

Figure 2 shows hierarchical clustering dendrogram for global COVID-19 data using the Ward linkage. The leaf nodes at the bottom of the figure are single countries, and the vertical links are the linkage distances at which the clusters join. Short linkages are between countries with similar epidemiological characteristics, and tall linkages are between more differing clusters. The tree structure highlights the nested clustering, illustrating how successively smaller sets of countries join up to merge into larger clusters until all nations are grouped in a single cluster at the top of the hierarchy. This representation shows information on the relative severity of pandemic similarity among countries and identifies the optimal number of clusters by identifying natural break points in the dendrogram.

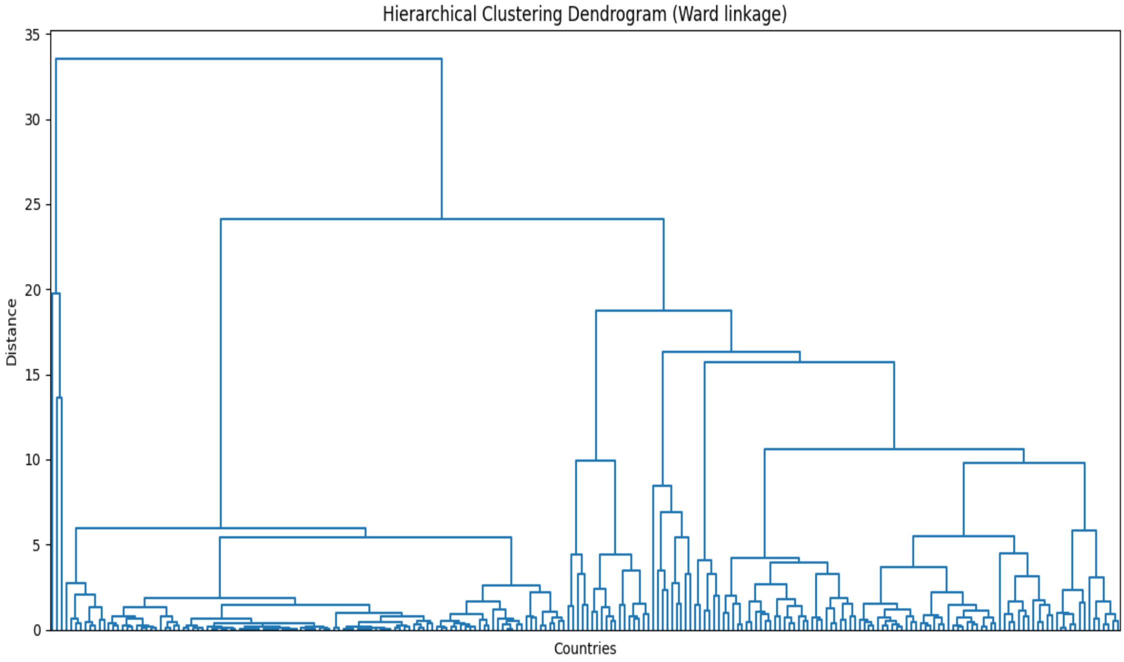


Figure 2: Hierarchical Clustering Dendrogram (Ward Linkage).

DOI: https://doi.org/10.54216/JAIM.100202
Received: April 02, 2025 Revised: June 28, 2025 Accepted: September 06, 2025

4. Result

The results of the research were made using Python, which provided an organized basis for statistical visualization and analysis. Libraries such as pandas, NumPy, and scikit-learn were used to implement data preprocessing, normalization, and clustering implementation, while matplotlib and seaborn were used to create dendrograms, correlation heatmaps, and cluster profile plots. This provided reproducible, accurate, and visually interpretable outcomes for hierarchical clustering analysis.

The PCA projection of the first two components of the data is presented in Figure 3. The four clusters are adequately represented by colour. Cluster 0 (50% of the countries) and Cluster 2 (45% of the countries) make up the bulk of the distribution, with Clusters 1 (~1%) and 3 (~1%) being far right extreme outliers in the plot.

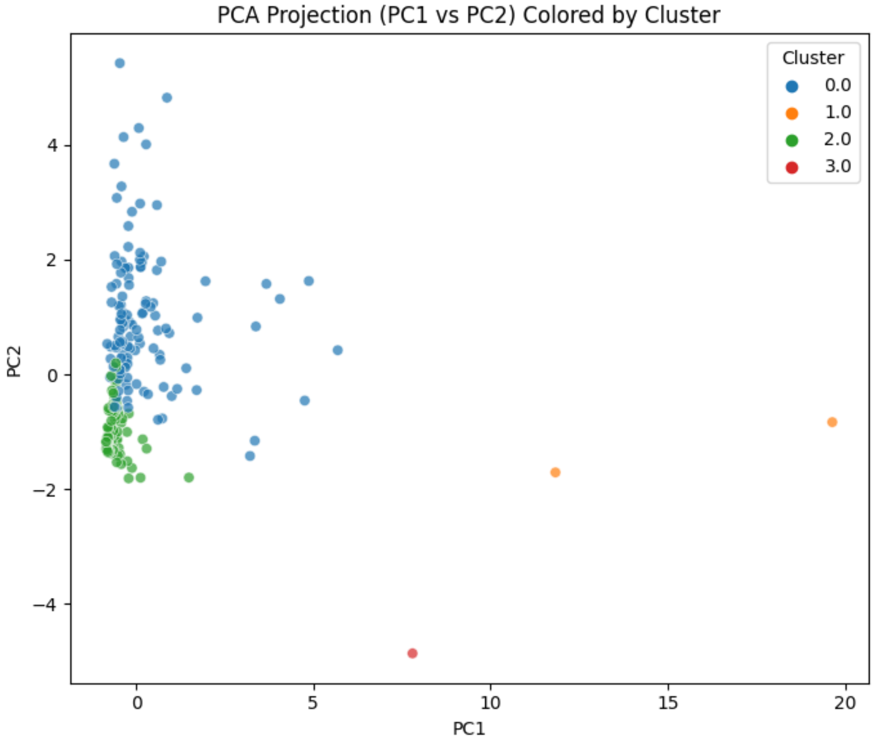


Figure 3: PCA Projection (PC1 vs PC2) Coloured by Cluster.

The first and third principal components are shown in Figure 4. Again, the majority of countries cluster in Cluster 0 (50%) and Cluster 2 (45%), with Clusters 1 and 3, containing less than 2% of the data set, clearly being outliers. This confirms the stability of the clustering across dimensions.

DOI: https://doi.org/10.54216/JAIM.100202
Received: April 02, 2025 Revised: June 28, 2025 Accepted: September 06, 2025

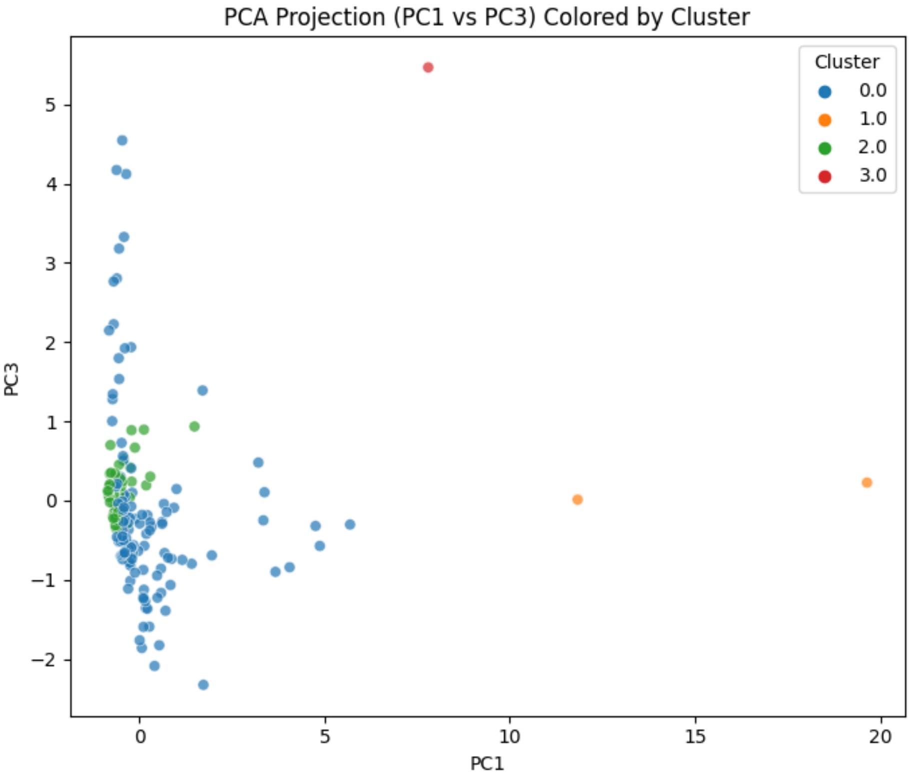


Figure 4: PCA Projection (PC1 vs PC3) Coloured by Cluster.

Figure 5 illustrates the agreement between principal components 2 and 3. Cluster 0 and Cluster 2 together include about 95% of all countries and form the dense middle of the plot. The ~5% remaining are isolated points in Clusters 1 and 3, once more evidence of rare pandemic profiles.

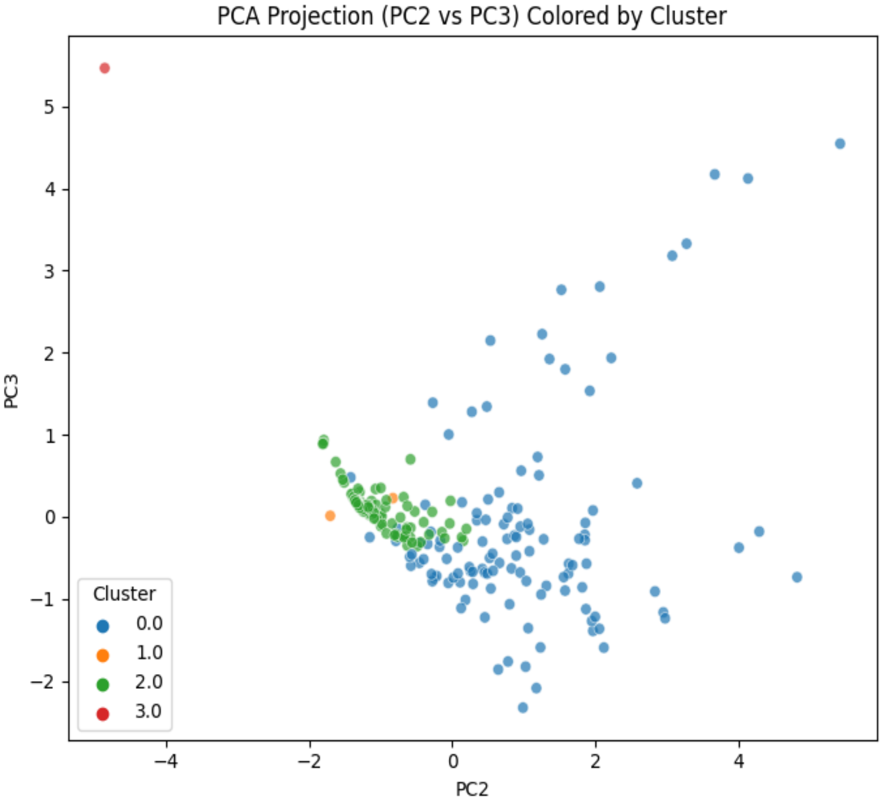


Figure 5: PCA Projection (PC2 vs PC3) Coloured by Cluster.

Figure 6 Presents the distribution of COVID-19 markers across the four clusters identified by hierarchical clustering. Each boxplot shows cluster variation and discloses central tendencies and outliers.

DOI: https://doi.org/10.54216/JAIM.100202
Received: April 02, 2025 Revised: June 28, 2025 Accepted: September 06, 2025

For Total Cases, Cluster 1 (~1% of countries) is the winner with values that are in astronomical range, ranging well into tens of millions. Cluster 0 (~50%) and Cluster 2 (~45%) contain the majority of the countries with moderate numbers, and Cluster 3 (~1%) captures intermediate but consistent figures.

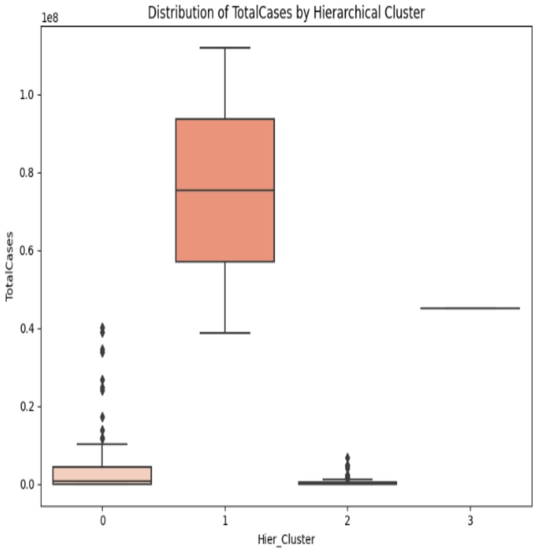
The Total Deaths distribution also captures this trend. Cluster 1, though extremely tiny at ~1% of nations, disproportionately represents world mortality, with over one million fatalities in certain cases. Cluster 0 and Cluster 2, with combined ~95% of nations, have relatively low median counts of deaths, capturing varied pandemic effect.

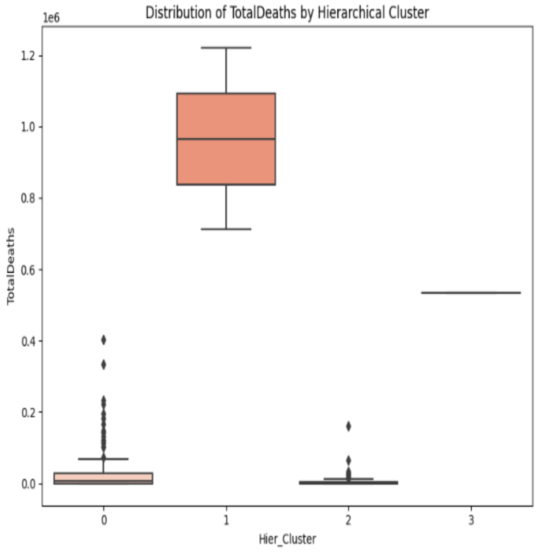
When considering Active Cases, Cluster 1 remains the most burdened with many countries possessing more than a million active cases. By contrast, Clusters 0 and 2, which account for a whopping 95% of all countries, possess much lower active loads, suggesting a dramatic difference between the world in general and some outlier, high-burden nations.

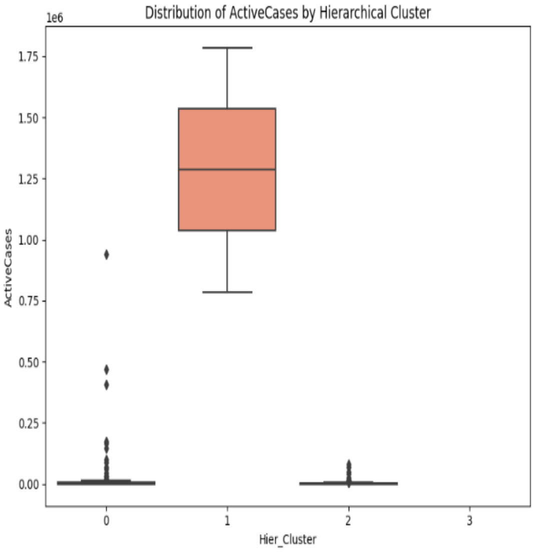
Per-capita data further supports this disparity. The Cases per Million subplot graphs Cluster 0 (~50%) and Cluster 2 (~45%) as having broad but tall ranges, as anticipated with widespread transmission. The Deaths per Million boxplot indicates that Cluster 1 nations (~1%) report the highest per-capita mortality rates, while Cluster 0 has a broader but more scattered distribution.

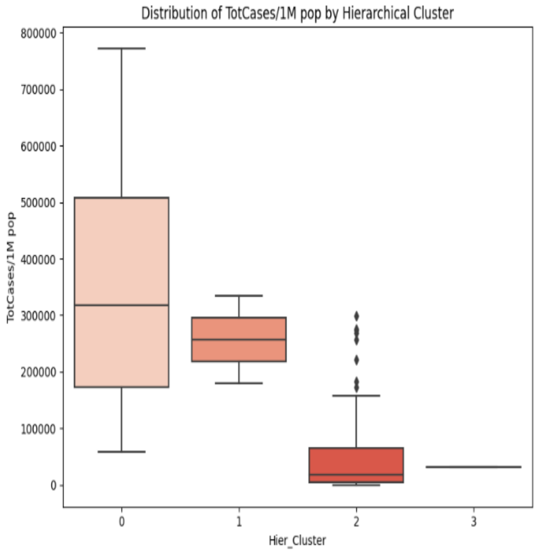
Intensity of testing, as reflected in the Tests per Million subplot, is most dense in Cluster 0, where certain nations have more than 20 million tests per million citizens. Cluster 2 reveals relatively moderate but consistent testing in ~45% of the data, and Clusters 1 and 3 remain underrepresented with relatively low coverage.

The Population subplot reveals striking demographic contrasts. Cluster 3, despite having merely ~1% of countries, is dominated by one immensely populated country, while most countries in Clusters 0 and 2 span moderate ranges.









DOI: https://doi.org/10.54216/JAIM.100202
Received: April 02, 2025 Revised: June 28, 2025 Accepted: September 06, 2025

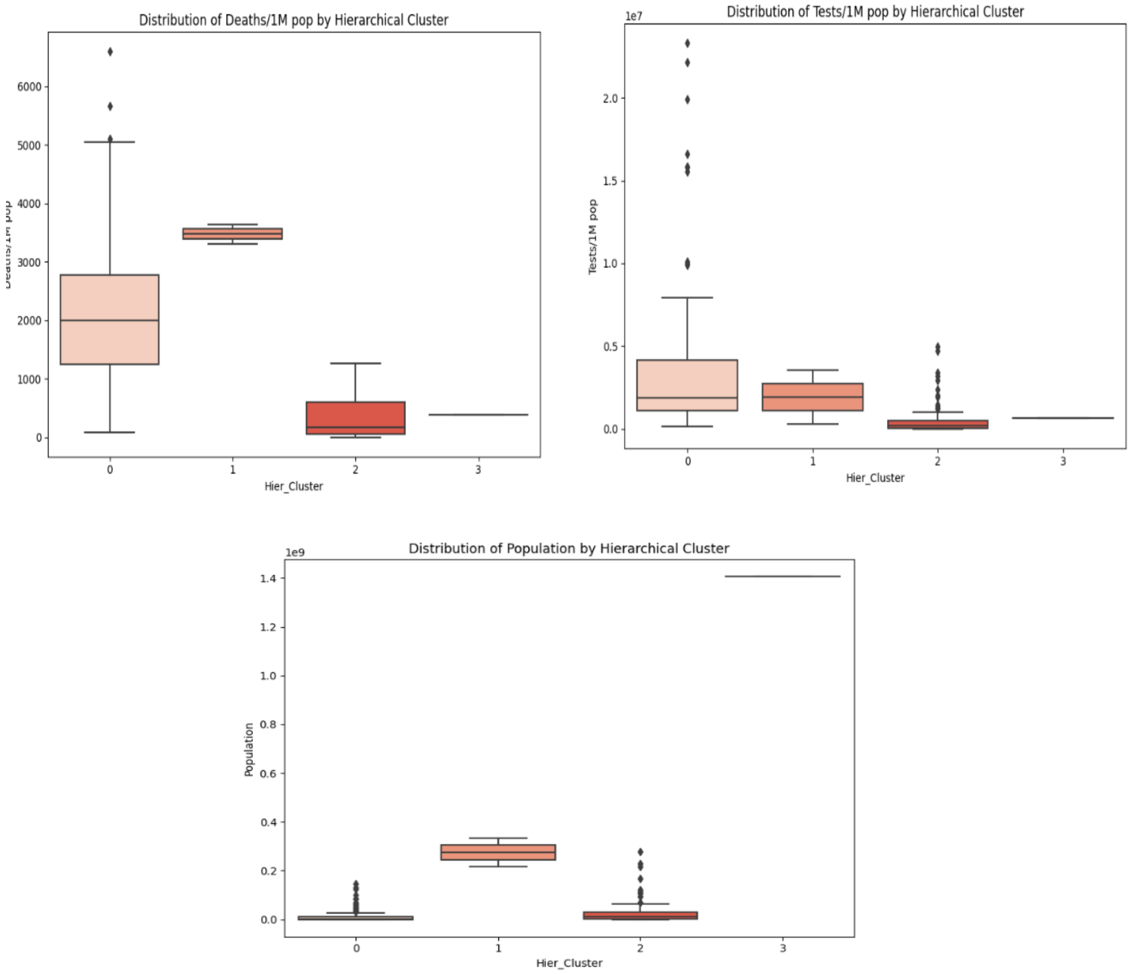


Figure 6: Distribution of COVID-19 Indicators by Hierarchical Cluster.

Figure 7 quantifies the cluster affiliation: Cluster 0 holds ~50% of countries, Cluster 2 holds ~45%, and Clusters 1 and 3 hold less than 2%. This shows that the global COVID-19 outbreak largely belongs to two primary epidemiologic patterns with few countries being substantially different.

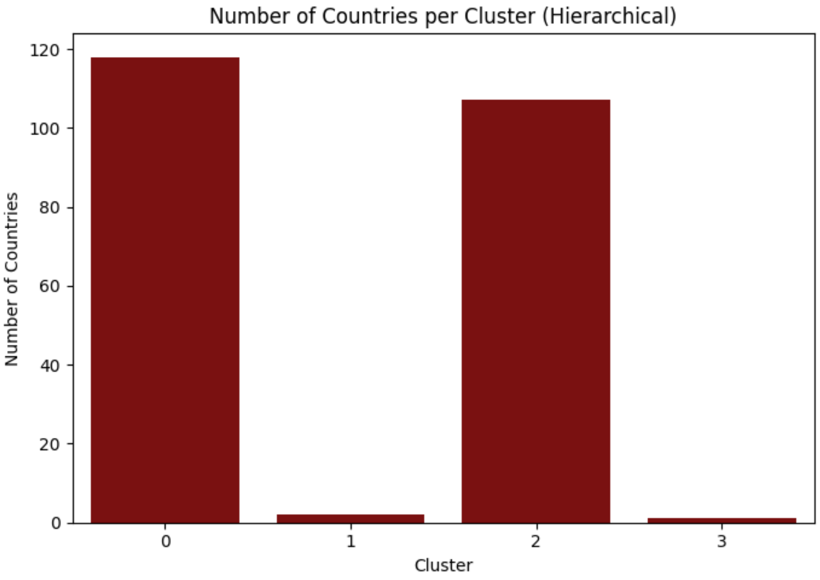


Figure 7: Number of Countries per Cluster.

DOI: https://doi.org/10.54216/JAIM.100202
Received: April 02, 2025 Revised: June 28, 2025 Accepted: September 06, 2025

Figure 8 illustrates the normalized mean values of COVID-19 indicators over the four hierarchical clusters. The heatmap indicates the relative contribution of each feature to the determination of the clusters, from 0 (minimum) to 1 (maximum).

Cluster 0 is characterized by extremely high per-capita levels of overall indicators, i.e., cases per million, deaths per million, and tests per million (all normalized to 1.0). Despite these per-capita intensities, absolute numbers (cases, deaths, recoveries) are small (~0.05), which suggests that this cluster is characteristic of smaller countries with relatively high pandemic impact when adjusted for population.

Cluster 1 bears the greatest burden in nearly all absolute metrics, with normalized counts of 1.0 for cases, deaths, recoveries, and active cases. Elevated measures of cases per million (0.72) and deaths per million (1.0) also indicate scale and severity. Although intensity of testing is moderate (0.46), this cluster represents the pandemic's global epicentres, where population size and burden were elevated.

Cluster 2 has very low normalized values for all indicators and the values are close to 0.0. Cluster 2 portrays nations with minimal reported pandemic impact, and the reason is they experience minimal case numbers, fatalities, and testing.

Cluster 3 displays a unique pattern in which both case (0.60) and death (0.55) counts are comparatively high while the highest normalized population is 1.0. This shows that the cluster represents high-population nations where absolute cases are high but per-capita impacts are contained due to the size of the base population.

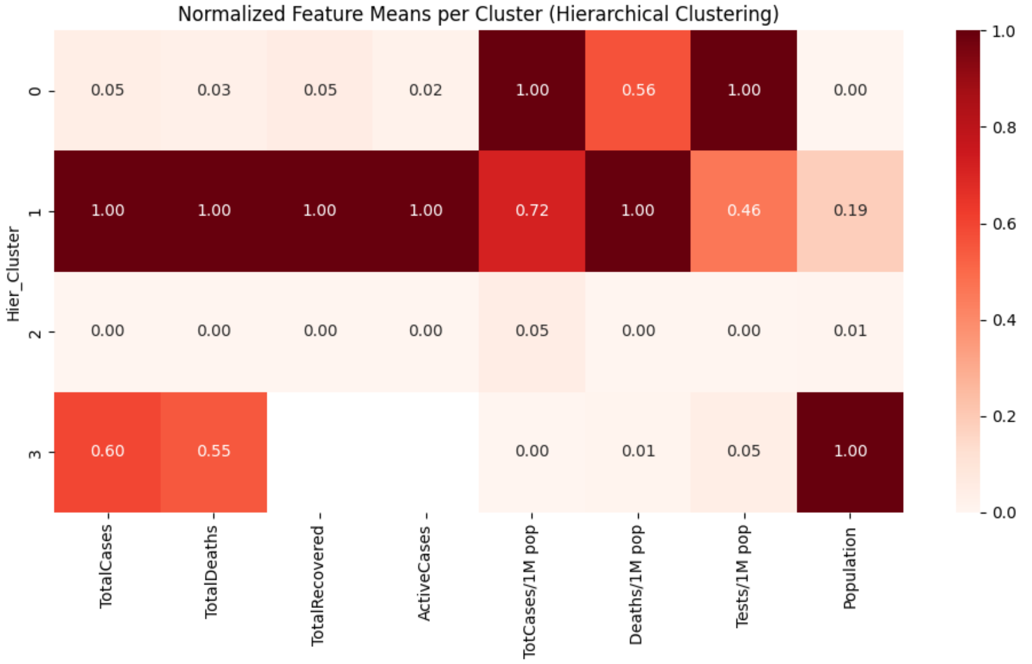


Figure 8. Normalized Feature Means per Cluster (Hierarchical Clustering)

5. Conclusion

This research used hierarchical clustering with Ward linkage to global COVID-19 data through January 2025 to identify patterns of pandemic effect by country. Four clusters were identified via thorough preprocessing, normalization, and clustering, representing various degrees of case burden, mortality, test intensity, and demographic characteristics.

Dendrogram revealed natural splits in the data, and PCA plots verified that the clusters were strongly separated. Cluster validity scores verified the quality of the clustering structure, as reflected by a Silhouette measure of 38.5%, a Davies–Bouldin Index of 0.87, and a Calinski–Harabasz Index of 77.6.

Cluster profiling found that two prominent clusters characterized the global dataset: Cluster 0, encompassing approximately 50% of countries, and Cluster 2, with approximately 45%, both under moderate to high-burden settings.

Clusters 1 and 3 (approximately 2%) comprised individual outlier countries with exceptionally high numbers of cases, outlier rates of testing intensity, or uncharacteristic demographic characteristics. Analysis of correlation indicated interdependencies among the variables, with high correlations between deaths, total cases, and recoveries, and per-capita aligned closely paralleling mortality rates.

DOI: https://doi.org/10.54216/JAIM.100202
Received: April 02, 2025 Revised: June 28, 2025 Accepted: September 06, 2025

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] World Health Organization, Joint external evaluation tool: international health regulations, 2018. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/259961/9789241550222-eng.pdf?sequence=1. [Accessed: January. 9, 2025].

[2] M. Moore, B. Gelfeld, A. Okunogbe, and C. Paul, “Identifying future disease hot spots: infectious disease vulnerability index,” Rand Health Q., vol. 6, no. 5, 2017. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/28845357

[3] M. Moore, B. Gelfeld, A. Okunogbe, and C. Paul, Identifying future disease hot spots: infectious disease vulnerability index. Santa Monica, CA: RAND Corporation, 2016.

[4] World Health Organization, “WHO Director-General’s opening remarks at the media briefing on COVID-19 – 11 March 2020.” [Online]. Available: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

[5] Worldometer, “Coronavirus worldwide graphs.” [Online]. Available: https://www.worldometers.info/coronavirus/worldwide-graphs/#total-deaths. [Accessed: January. 9, 2025].

[6] T. Hale, N. Angrist, R. Goldszmidt, et al., “A global panel database of pandemic policies (Oxford COVID-19 government response tracker),” Nat. Hum. Behav., vol. 5, pp. 529–538, 2021, doi: 10.1038/s41562-021-01079-8.

[7] PreventEpidemics, “Joint External Evaluation (JEE) scores.” [Online]. Available: https://preventepidemics.org/wp-content/uploads/excel/all-countries.xlsx. [Accessed: January. 9, 2025].

[8] J. H. Ward, “Hierarchical grouping to optimize an objective function,” J. Am. Stat. Assoc., vol. 58, pp. 236–244, 1963, doi: 10.1080/01621459.1963.10500845.

[9] D. T. Jamison, L. J. Lau, K. B. Wu, et al., “Country performance against COVID-19: rankings for 35 countries,” BMJ Glob. Health, vol. 5, e003047, 2020, doi: 10.1136/bmjgh-2020-003047.

[10] S. Kumar, “Use of cluster analysis to monitor novel coronavirus-19 infections in Maharashtra, India,” Indian J. Med. Sci., vol. 72, pp. 44–48, 2020, doi: 10.25259/IJMS_68_2020.

[11] P. Sengupta, B. G. Ghosh, and S. SenRoy, “An analysis of COVID-19 clusters in India – two case studies on Nizamuddin and Dharavi,” Research Square, 2020, doi: 10.21203/rs.3.rs-68814/v1.

[12] M. L. F. Nascimento, “A multivariate analysis on spatiotemporal evolution of COVID-19 in Brazil,” Infect. Dis. Model., vol. 5, pp. 670–680, 2020, doi: 10.1016/j.idm.2020.08.012.

[13] “K-means clustering of COVID-19 cases in Indonesia’s Provinces,” Proc. Int. Conf. Global Optimization and Its Applications, Nov. 21–22, 2020.

[14] N. James and M. Menzies, “COVID-19 in the United States: trajectories and second surge behavior,” Chaos, vol. 30, p. 091102, 2020, doi: 10.1063/5.0024204.

[15] V. Zarikas, S. G. Poulopoulos, Z. Gareiou, et al., “Clustering analysis of countries using the COVID-19 cases dataset,” Data Brief, vol. 31, p. 105787, 2020, doi: 10.1016/j.dib.2020.105787.

[16] M. R. Mahmoudi, D. Baleanu, Z. Mansor, et al., “Fuzzy clustering method to compare the spread rate of COVID-19 in the high risks countries,” Chaos Solitons Fractals, vol. 140, p. 110230, 2020, doi: 10.1016/j.chaos.2020.110230.

[17] O. Pasin, “Clustering of countries in terms of deaths and cases of COVID-19,” J. Health Soc. Sci., vol. 5, pp. 587–594, 2020.

[18] A. Ramadan, A. Kamel, A. Taha, et al., “A multivariate data analysis approach for investigating daily statistics of countries affected with COVID-19 pandemic,” Heliyon, vol. 6, p. e05575, 2020, doi: 10.1016/j.heliyon. 2020.e05575.

[19] N. James and M. Menzies, "Cluster-based dual evolution for multivariate time series: analyzing COVID-19," Chaos, vol. 30, p. 061108, 2020, doi: 10.1063/5.0013156.

[20] M. M. Kavanagh, R. Singh, and S. R. Democracy, "Democracy, capacity, and coercion in pandemic response: COVID-19 in comparative political perspective," J. Health Polit. Policy Law, vol. 45, pp. 997–1012, 2020, doi: 10.1215/03616878-8641530.

[21] Global Health Security (GHS) Index. [Online]. Available: http://www.ghsindex.org. [Accessed: Aug. 9, 2021].

[22] B. Oppenheim, M. Gallivan, N. K. Madhav, et al., "Assessing global preparedness for the next pandemic: development and application of an epidemic preparedness index," BMJ Glob. Health, vol. 4, p. e001157, 2019, doi: 10.1136/bmjgh-2018-001157.

[23] W. J. Wiersinga, A. Rhodes, A. C. Cheng, et al., "Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review," JAMA, vol. 324, pp. 782–793, 2020, doi: 10.1001/jama.2020.12839.

[24] S. Richardson, J. S. Hirsch, J. Narasimhan, et al., "Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area," JAMA, vol. 323, pp. 2052–2059, 2020, doi: 10.1001/jama.2020.6775.

[25] D. M. Weinberger, J. Chen, T. Cohen, et al., "Estimation of excess deaths associated with the COVID-19 pandemic in the United States, March to May 2020," JAMA Intern. Med., vol. 180, pp. 1336–1344, 2020, doi: 10.1001/jamainternmed.2020.3391.

[26] J. Dumlao, "Global COVID-19 Statistics (Jan 2025)," Kaggle, 2025. [Online]. Available: https://www.kaggle.com/datasets/jocelyndumlao/global-covid-19-statistics-jan-2025.

[27] Alhasani T. A., Alkattan H., Ali A., El-Kenawy E.-S. M., and Eid M. M., "A Comparative Analysis of Methods for Detecting and Diagnosing Breast Cancer Based on Data Mining," Journal of Artificial Intelligence and Metaheuristics, vol. 4, no. 2, pp. 08-17, 2023, doi: 10.54216/JAIM.040201.

[28] Doaa S. Khafaga, Hussein Alkattan, Alhumaima A. Subhi, "Evaluating the Effect of Optimized Voting Using Hybrid Particle Swarm and Grey Wolf Algorithm on the Classification of the Zoo Dataset," Journal of Artificial Intelligence and Metaheuristics, vol. 2, no. 1, pp. 08-15, 2022, doi: 10.54216/JAIM.020101.

[29] H. K. Al-Mahdawi, M. Abotaleb, H. Alkattan, E.-S. M. El-Kenawy, and E. M. Mohamed, "Solving the Inverse Initial Value Problem for the Heat Conductivity Equation by Using the Picard Method," Journal of Artificial Intelligence and Metaheuristics, vol. 2, no. 2, pp. 46-55, 2022, doi: 10.54216/JAIM.020205.

[30] H. Alkattan, S. K. Towfek, and M. Y. Shams, "Tapping into Knowledge: Ontological Data Mining Approach for Detecting Cardiovascular Disease Risk Causes Among Diabetes Patients," Journal of Artificial Intelligence and Metaheuristics, vol. 4, no. 1, pp. 08-15, 2023, doi: 10.54216/JAIM.040101.

[31] Noor Razzaq Abbas, Hussein Alkattan, Hamidreza Rabiei-Dastjerdi, Mohamed Saber, and Marwa M. Eid, "Monthly Solar Prediction Using Machine Learning: Diyala Governorate, Iraq as a Case Study," Journal of Artificial Intelligence and Metaheuristics, vol. 5, no. 2, pp. 41-46, 2023, doi: 10.54216/JAIM.050204.

[32] Al-Mahdawi H. K., Alhumaima Ali Subhi, Hussein Alkattan, Mohamed Saber, Marwa M. Eid, Anfal A. Sabti Al-Mahdawi, and Jinan A. M. Al-Saddaee, "Solving Initial Value Problem in Composite Materials for Heat Equation," Journal of Artificial Intelligence and Metaheuristics, vol. 6, no. 1, pp. 08-17, 2023, doi: 10.54216/JAIM.060101.

[33] Al-Seyday, T. Qenawy; Hussein Alkattan; Amany Khaled, "Securing DNS over HTTPS: A Machine Learning Study on Traffic Classification Using DoHBrw-2020," Journal of Artificial Intelligence and Metaheuristics, vol. 7, no. 2, pp. 73-81, 2024, doi:10.54216/JAIM.070207.

[34] A. Gupta and R. Kumar, "A Hybrid Deep Learning Approach for Brain Stroke Detection Using MRI Images," Journal of Medical Systems, vol. 46, no. 3, pp. 1-10, 2022. doi: 10.1007s10916-022-01844-7.

DOI: https://doi.org/10.54216/JAIM.100202
Received: April 02, 2025 Revised: June 28, 2025 Accepted: September 06, 2025